# Search Index (Earlybird) core classes

> \*\*TL;DR\*\* Earlybird (Search Index) find tweets from people you follow, rank them, and serve tweets to Home.

## What is Earlybird (Search Index)

[Earlybird](http://notes.stephenholiday.com/Earlybird.pdf) is a \*\*real-time search system\*\* based on [Apache Lucene](https://lucene.apache.org/) to support the high volume of queries and content updates. The major use cases are Relevance Search (specifically, Text search) and Timeline In-network Tweet retrieval (or UserID based search). It is designed to enable the efficient indexing and querying of billions of tweets, and to provide low-latency search results, even with heavy query loads.

## Directory Structure

The project consists of several packages and files, which can be summarized as follows:

\* `facets/`: This subdirectory contains classes responsible for facet counting and processing. Some key classes include EarlybirdFacets, EarlybirdFacetsFactory, FacetAccumulator, and FacetCountAggregator. The classes handle facet counting, facet iterators, facet label providers, and facet response rewriting.

\* `index/`: This directory contains the indexing and search infra files, with several subdirectories for specific components.

\* `column/`: This subdirectory contains classes related to column-stride field indexes, including ColumnStrideByteIndex, ColumnStrideIntIndex, ColumnStrideLongIndex, and various optimized versions of these indexes. These classes deal with managing and updating doc values.

\* `extensions/`: This subdirectory contains classes for index extensions, including EarlybirdIndexExtensionsData, EarlybirdIndexExtensionsFactory, and EarlybirdRealtimeIndexExtensionsData.

\* `inverted/`: This subdirectory focuses on the inverted index and its components, such as InMemoryFields, IndexOptimizer, InvertedIndex, and InvertedRealtimeIndex. It also contains classes for managing and processing posting lists and term dictionaries, like EarlybirdPostingsEnum, FSTTermDictionary, and MPHTermDictionary.

\* `util/`: This subdirectory contains utility classes for managing search iterators and filters, such as AllDocsIterator, RangeDISI, RangeFilterDISI, and SearchSortUtils. The system appears to be designed to handle search indexing and facet counting efficiently. Key components include an inverted index, various types of posting lists, and term dictionaries. Facet counting and processing is handled by specialized classes within the facets subdirectory. The overall structure indicates a well-organized and modular search indexing system that can be maintained and extended as needed.

## Related Services

\* The Earlybirds main classes. See `src/java/com/twitter/search/earlybird/`